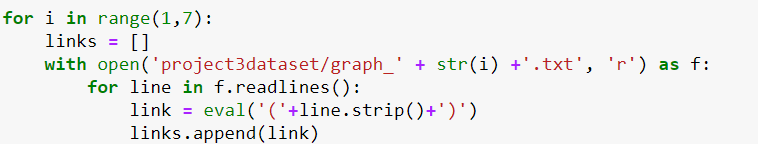
Link Analysis Practice

E94056178 工科系 盧柏翔

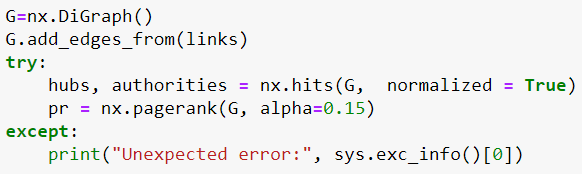
* **Implementation(詳見main.ipynb)**

Task 1: Please implement HITS and PageRank

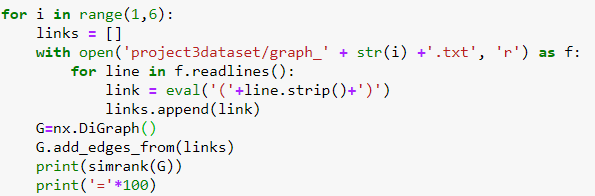
STEP1:讀檔並建立連結存至list



STEP2-1:利用python的networkx套件分別計算各Graph的hub、authority、pagerank值



STEP2-2:計算前5 Graph的SimRank值

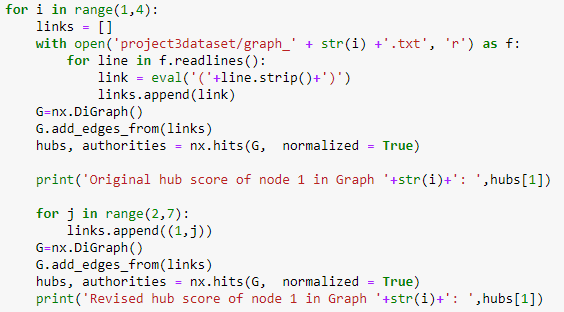


SimRank 函式實作

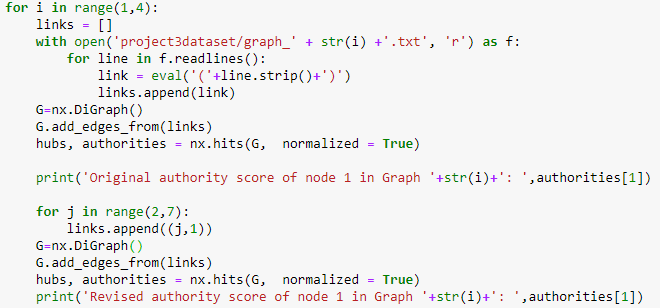


Task 2: Find a way to increase hub, authority and PageRank of Node 1 in first 3 graphs respectively

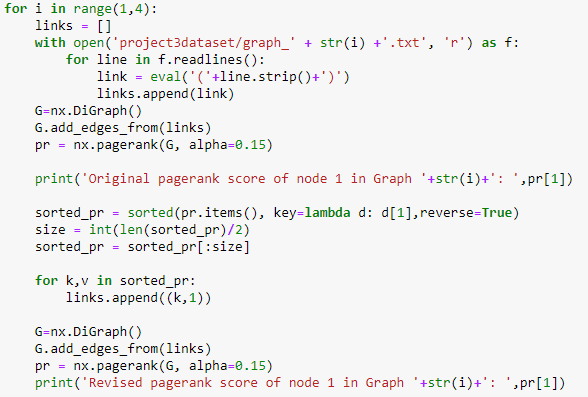
1. 提高Node1 hub值(想法:增加Node 1到其他點的link)



1. 提高Node1 authority值(想法：增加其他點到Node 1的link)

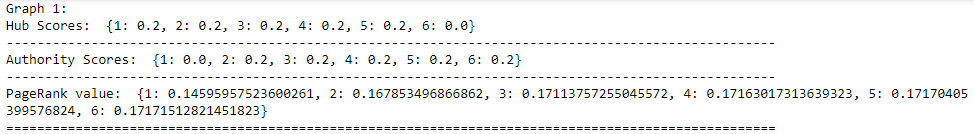
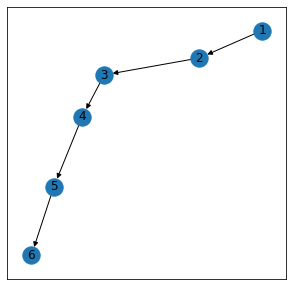


1. 提高PageRank 值(想法：盡量找PR值比較高且外部連結又少的點)



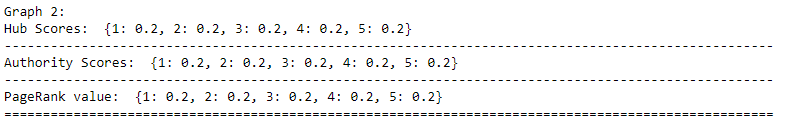
* **Result analysis and discussion**

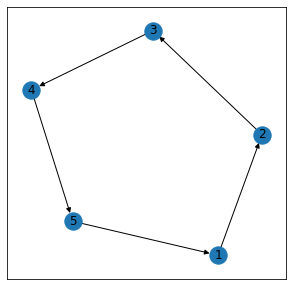
**各graph的hub,authority 和 pagerank的值**

1. **Graph 1**

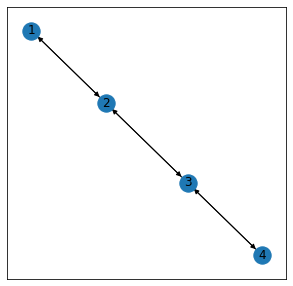
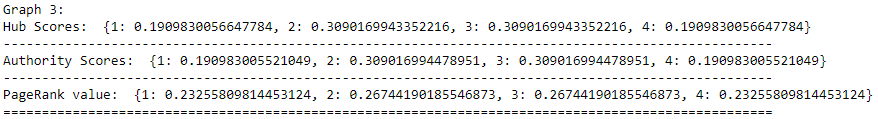
分析：Graph1 圖為單向一直線，可以注意到在Node1因無parent node所以authority為0，Node6因無chlid node 所以hub為0，使用PageRank則可以避免此現象，因演算法賦予每個頁面最小值。

****

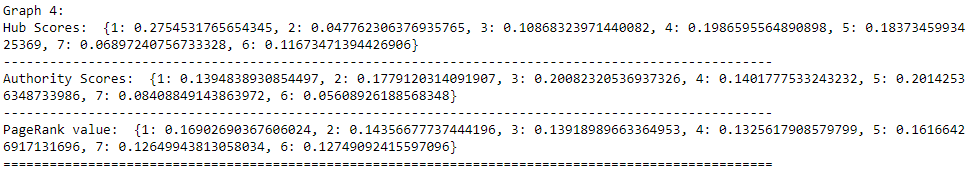
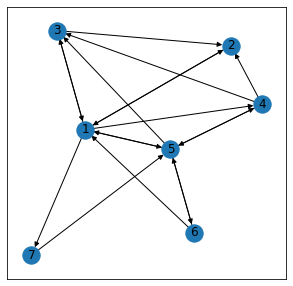
1. **Graph 2**



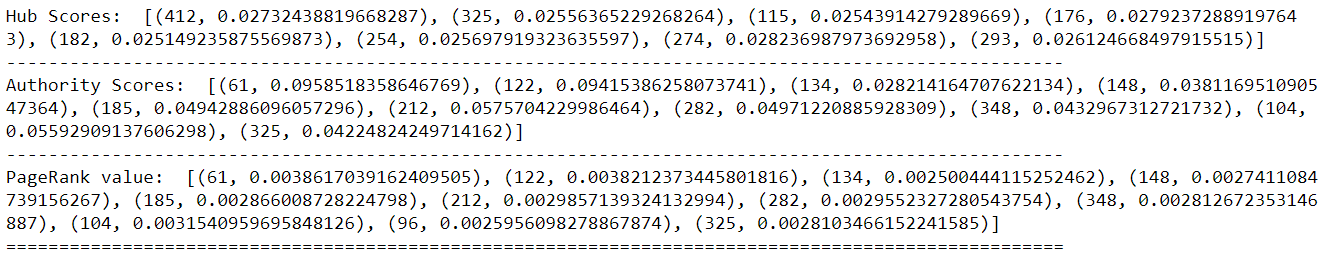
分析：Graph2 圖為單向環狀，導致hub,authority和pagerank值的結果跟初始值一樣並未改變。Hub值為child的authority相加再歸一化，Authority值為parent的hub相加再歸一化，pagerank則像水流般傳值，每個點的前後狀況均相同，因此值均一樣。

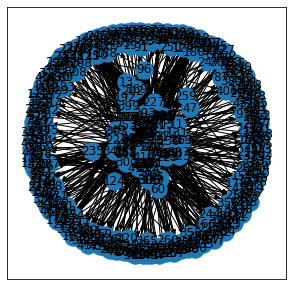
1. **Graph 3**

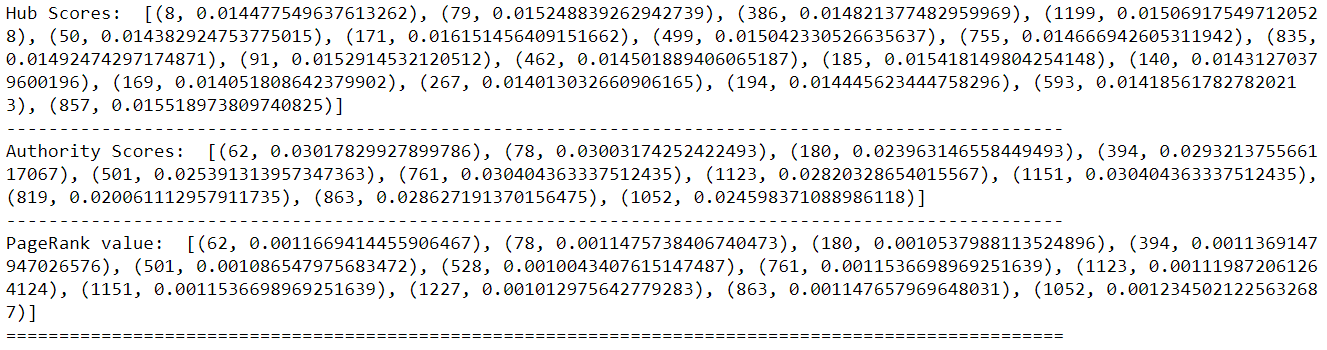
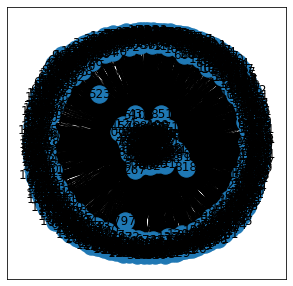
分析：Graph 3圖為直線雙向連結，

1. **Graph 4**

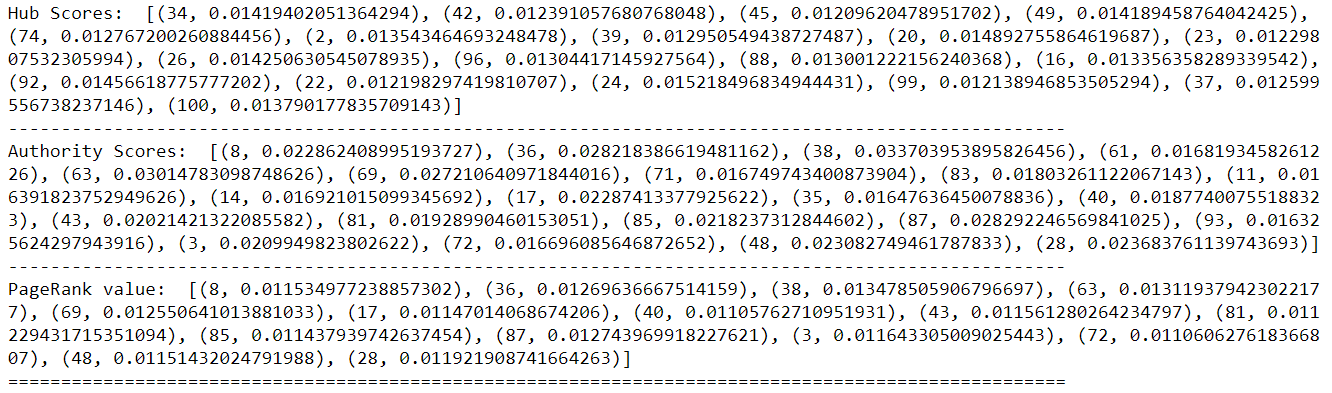
分析：Graph 4圖為

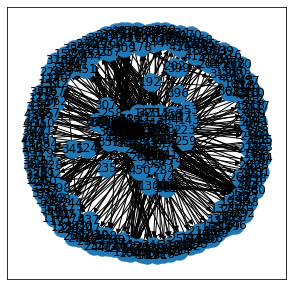
1. **Graph 5**

分析：此graph較複雜，有需多node但是大部分的node並沒有太多的Indegree和Outdegree，在weights初始均分的情況下，大部份的值都在重要少數的node上，造成其他node的hub,authrotiy,pagerank都趨近為零。此處列出前幾個值較大的node，發現hub,authority,pagerank的node都差不多。

1. **Graph 6**

分析：此graph最複雜，約進行了120次iteration才達到收斂，收斂的結果跟graph5情況相似，大部分的node的hub,authority,pagerank都趨近零。此處列出前幾個值較大的node，發現hub,authority,pagerank的node都差不多。

1. **Graph from transaction data**

分析：由transaction data產生的graph，大部分的node未與其他node連接，因此跟前兩個graph相似，node多導致值小且集中在少數node上。

**SimRank**

**增加hub, authority和pagerank**

* Computation performance analysis
* Discussion

More limitations about link analysis algorithms

Answer:

Can link analysis algorithms really find the “important”

pages from Web?

Answer:

What are practical issues when implement these

algorithms in a real Web?

Answer: 實際各個網頁是動態的，不容易收斂變穩定，且太多的廣告連結會干擾分析效果。

What do the result say for your actor/movie graph?

Any new idea about the link analysis algorithm?

Answer:

What is the effect of “C” parameter in SimRank?

Answer:

Design a new link-based similarity measurement

Answer: